Available online at www.sciencedirect.com
SCIENCE DIRECT
Vision Research 46 (2006) 1911–1923
Vision Research
www.elsevier.com/locate/visres

Stereo dynamics are not scale-dependent

Robert F. Hess *, Laurie M. Wilcox

McGill Vision Research, Department of Ophthalmology, McGill University, Montreal, Que., Canada
Centre for Vision Research, Department of Psychology, York University, Toronto, Ont., Canada

Received 15 April 2005; received in revised form 17 November 2005

Abstract

The experiments reported here focus on the temporal dynamics of stereopsis in an effort to shed light on how low level mechanisms might contribute to the execution of coarse-to-fine processing in the human stereo system. Because previous studies have used a variety of stimuli and configurations, we assess the effect of exposure duration on stereo thresholds using band-limited Gabor patches for a range of stimulus configurations. In preliminary studies, we found that the best stereo sensitivity–spatial frequency relationship was obtained when using configurations in which the size and target–reference spacing were consistent with spatially *scaled* stimuli. Sub-optimal stereo sensitivity as a function of spatial frequency was observed when the size and separation were fixed. Further, we found that the temporal properties of stereopsis were consistently sustained in nature irrespective of the stimulus spatial frequency content. This latter finding suggests that if coarse-to-fine stereo processing does occur it does not follow as a consequence of the dynamics of low-level disparity transduction.

Keywords: Stereopsis; Spatial frequency; Dynamics; Spatial properties

1. Introduction

Recent neurophysiological evidence suggests that there is a strong temporal dependence for stereoscopic processing (Menz & Freeman, 2003, 2004a, 2004b). These studies showed that the cortex processed disparity information at coarse scales faster than it processed disparity signals at fine scales. This in turn suggests that the coarse disparities that are conveyed by low spatial frequency neurons have faster dynamics than fine disparities that are conveyed by higher spatial frequency neurons. Such a finding fits well with the well known fact that low spatial frequency stimuli are detected more rapidly than high spatial frequency stimuli (Burr, 1981). Furthermore, it offers a convenient low-level explanation for coarse-to-fine sequential processing of stereoscopic information (Mallot, Gillner, & Arndt, 1996; Rohaly & Wilson, 1993, 1994; Smallman & MacLeod, 1994; Watt, 1987), a processing strategy which some (Marr & Poggio, 1979) have suggested would help solve inherent ambiguities associated with stereo processing of broadband images.

Apart from the above neurophysiology there is little empirical evidence that addresses whether stereo dynamics vary with stimulus spatial scale, however, a recent study of stereo processing in human and non-human primates, suggests it does not. That is, Harwerth, Fredenburg, and Smith (2003) show that there is no change in the dynamics of stereo discrimination as a function of spatial frequency. In their study however, they use stimuli that have a fixed separation and spatial frequency bandwidth but variable orientation bandwidth. It is not clear the extent to which their particular choice of stimulus geometry was crucial to this finding. Since this is such a critical finding we wanted to re-investigate the issue of whether stereo dynamics vary with stimulus spatial scale and in so doing use a variety of stimulus configurations in case stimulus geometry is important. These conditions include stimuli in which spatial frequency is

* Corresponding author.
E-mail address: robert.hess@mcgill.ca (R.F. Hess).

0042-6989/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.12.007

varied but where the individual elements are of fixed size and separation (Schor, Edwards, & Pope, 1998), fixed size but variable separation, variable size but fixed separation (e.g., Harwerth et al., 2003) and finally of scaled size and separation (e.g., Hess & Wilcox, 1994). Since our aim was not to provide definitive data on the relationship between stereo performance and the parameters of size, bandwidth and separation, we selected a range of stimulus geometries that were representative of what has been used in previous stereo experiments. This enabled us to assess whether stimulus geometry played a crucial role in existing reports on how stereo dynamics vary with spatial frequency. In the first part of the study, we compare how stereo acuity varies with spatial frequency for different commonly used stimulus configurations in which element separation, bandwidth and envelope size vary. In the second part of the study, we compare the dynamics as a function of stimulus spatial frequency by modeling the relationship between stereo sensitivity and exposure duration and we assess the degree to which the derived model parameter varies with different stimulus configurations.

2. Methods

2.1. Apparatus

Stimuli were presented as grey level variations on a Clinton monitor. A full screen display of 1024×768 pixels was used. At a viewing distance of 1.15 m this subtended $17° \times 14°$ of visual angle. The mean luminance was 69 cd/m^2 and the screen remained at mean luminance except when stimuli were presented. The monitor was controlled by a Cambridge Research Systems VSG2/3 graphics card which implements a resistor network to sum DAC outputs and allows a pseudo 12-bit representation of grey after gamma correction. The frame rate was 120 Hz. Stereo pairs were displayed on alternate frames and seen by each eye using LCD goggles synchronized to the frame rate.

2.2. Observers

Five subjects (including one of the authors) were tested. Each of the observers had normal or corrected to normal vision with normal stereovision as assessed using the Randot Stereotest and by their performance in previous stereo experiments.

2.3. Stimuli

Stimuli were presented as arrangements of Gabor patterns. Each of these had a luminance distribution of the form:

$$L(x,y) = L_0\{1 + C \cdot \cos(2\pi/\lambda \cdot (x\cos\theta + y\sin\theta)) \cdot \exp((x^2+y^2)/\sigma^2)\}, \quad (1)$$

where L_0 is the mean luminance, C is the contrast of the Gabor, λ is the period of the carrier, θ is the orientation term, x and y are the distances along horizontal and vertical axes, respectively (from the Gabor centre) and σ the standard deviation of the Gaussian envelope.

In these experiments the Gabor spatial frequency, standard deviation and inter-Gabor spacing was varied. Examples of the stimuli are shown in Fig. 1.

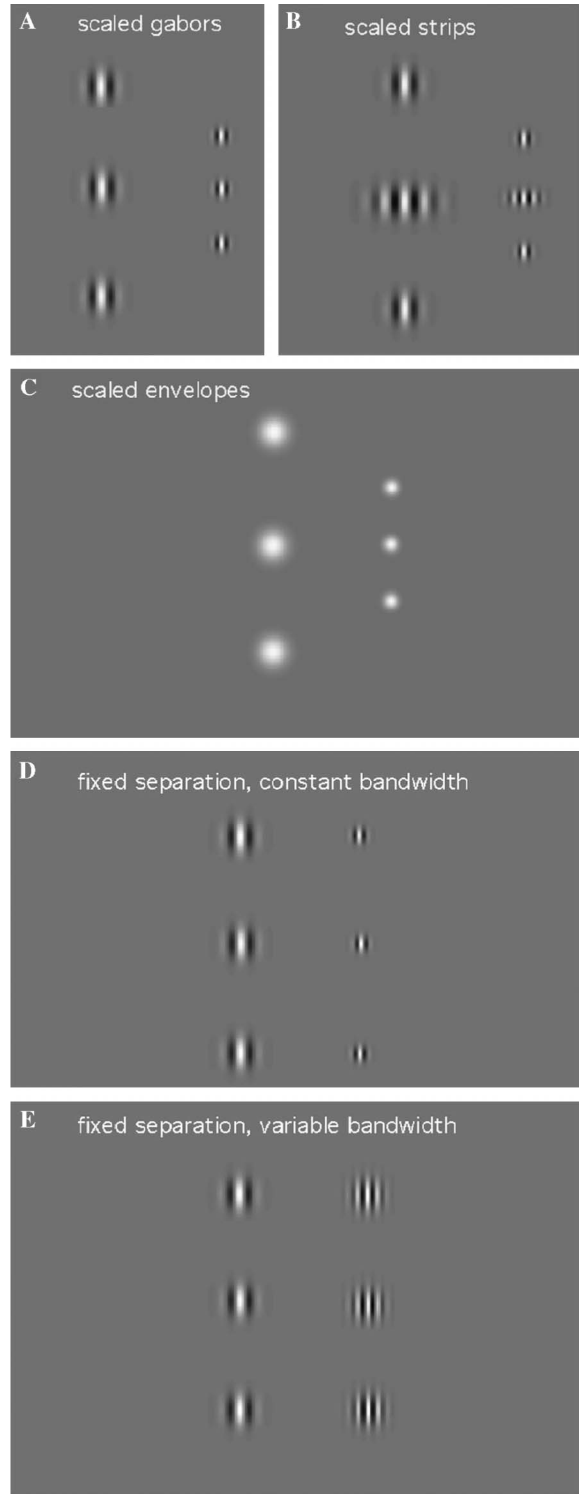

Fig. 1. Subjects were required to judge whether the central element was in front or behind the outer two, zero disparity reference elements. Three types of scaled stimuli used to measure stereo sensitivity as a function of spatial frequency/spatial scale in the current experiment. In (A), all aspects of the stimuli are scaled with spatial frequency (scaled Gabors), in (B), the central target is 2.5 times wider horizontally than vertically but all aspects were scaled with spatial frequency (scaled strips) and in (C), envelope separation and size were spatially scaled (scaled envelopes). In (D), stereo stimuli had a fixed separation and constant bandwidth as spatial frequency was varied. In (E), stimuli had a fixed separation but a variable bandwidth as a function of spatial frequency.

The standard stimulus arrangement consisted of a Gabor triplet in which the two outer Gabors served as reference elements and were always at zero disparity. The middle Gabor was the target and was positioned, randomly, at one of a number of crossed and uncrossed disparities. The test and reference stimuli were always spatially identical except in the scaled strip configuration (Fig. 1B). Over a series of test sessions (across several months) stereo sensitivity was measured using a range of commonly used stimulus parameters (see Table 1) to allow assessment of the individual effects of exposure duration, as a function of spatial frequency, envelope size (or bandwidth), and stimulus/reference separation. Exposure duration varied within conditions from a minimum of 122 ms to a maximum of 1336 ms.

When measuring stereoscopic thresholds as a function of exposure duration it is important to limit the visible interval as precisely as possible. While it is impossible to be sure that stimuli are not processed after they are extinguished, the likelihood of post-stimulus processing is greatly reduced by using an appropriate post-stimulus mask (Breitmeyer & Ogmen, 2000; Kahneman, 1968; Liss, 1968; Sperling, 1963). Accordingly, the mask pattern used here was a horizontal strip of luminance noise, created by assigning each vertical column of pixels a random luminance value. On each trial a different, randomly selected, noise pattern was presented to each eye, creating the percept of a volume of disparate bars. The mask was equivalent to the test in Michelson contrast and was wide enough to completely cover the central test stimulus. For all conditions the mask appeared immediately following the test stimulus, and was visible for 50 ms. Previous studies have shown that decisions based on stereoscopic information beyond this time period are ineffective (Julesz, 1964; Uttal, Fitzgerald, & Eskin, 1975).

Table 1
Parameter values

SF (c/d)	σ_x (min)	σ_y (min)	Separation (min)
A. Scaled Gabors			
0.4	54	54	135
0.8	30	30	75
1.6	15	15	37.5
3.2	7.5	7.5	18
6.4	3.6	3.6	9
B. Scaled strips			
0.4	135	54	135
0.8	75	30	75
1.6	37.5	15	37.5
3.2	18	7.5	18
6.4	9	3.6	9
C. Scaled envelopes			
0.08	54	54	135
0.08	30	30	75
0.08	15	15	37.5
0.08	7.5	7.5	18
0.08	3.6	3.6	9
D. Gabors of fixed separation and fixed bandwidth			
0.4	54	54	150
0.8	30	30	150
1.6	15	15	150
3.2	7.5	7.5	150
6.4	3.6	3.6	150
E. Gabors of fixed separation and variable bandwidth			
0.4	30	30	150
0.8	30	30	150
1.6	30	30	150
3.2	30	30	150
6.4	30	30	150

3. Procedure

3.1. Contrast thresholds

To ensure that stimulus detectability would not limit stereo sensitivity (particularly for brief presentations) we presented all stimuli at 3.5 times their contrast threshold at each exposure duration for each stimulus configuration. Since absolute contrast has been shown to not affect the temporal properties of stereopsis (Harwerth et al., 2003; Ogle & Weil, 1958) equating the stimuli for detectability will not bias thresholds. Contrast thresholds were measured prior to stereo testing for each duration, stimulus configuration, and target type. For all contrast measurements, we used the method of adjustment with a randomized starting point to obtain 7 binocular threshold estimates, which were then averaged. In all cases, contrast thresholds were measured using the post-stimulus mask to ensure that the effect of the mask on visibility was taken into account. Once these thresholds were obtained, the reference and target stimuli used for subsequent estimation of stereo thresholds were set to 3.5 times this value. This ensured that all test stimuli were equated in terms of multiples of detection threshold for each subject and therefore any effects of duration, size or frequency were not due to changes in a given individual's absolute sensitivity for detecting the stimulus.

3.2. Stereo

Stereo sensitivity was measured using the method of constant stimuli, with a set of 11 stimuli that covered a range of crossed and uncrossed disparities. This range was chosen individually for each stimulus condition to bracket the point at which the perceived location of the central Gabor changed from being 'in front' to 'behind' the outer reference Gabors. When required, sub-pixel spatial accuracy was achieved by recomputing each newly located stimulus. The stimuli were presented within a temporal Gaussian window; the sigma of this Gaussian controlled the exposure duration. Stimulus presentation was immediately followed by a spatial mask consisting of vertical 1-D spatial noise to ensure processing of stereo information was limited to the presentation duration used in a given session. The observers' task was to identify on each trial whether the central target was positioned in front of or behind the two outer stimuli and within a single run each of the depth offsets were presented a minimum of 40 times in random order. A stereo sensitivity estimate was derived from the resulting psychometric function, by fitting the error function (cumulative normal), ERF(x), of the form:

$$P(x) = A(0.5 + 0.5\mathrm{erf}((x - B)/(\sqrt{2C}))), \quad (2)$$

where A is the number of presentations per stimulus condition, B is the offset of the function relative to zero, and C is the standard deviation of the assumed underlying, normally distributed, error function. We use the reciprocal of the standard deviation parameter as our measure of stereo sensitivity. The reported 95% confidence limits for the thresholds were derived using a bootstrap procedure (Wichmann & Hill, 2001) with 10,000 replications.

4. Results

As noted above, our initial aim is to determine how sensitivity for stereo processing varied with stimulus spatial frequency for a selection of stimulus configurations that are representative of what has been used previously. Having done that we next wanted to investigate how the *dynamics* of stereo processing vary with stimulus spatial frequency and, by using a variety of stimulus configurations, gauge whether stimulus configuration is a crucial factor when considering the dynamics. With this dual goal in mind, we will begin by focusing on the spatial attributes of the stimuli, and their influence on *sensitivity*, followed by examination of the *dynamics*.

The spatially scaled stimulus, shown in Fig. 1A, is the simplest and perhaps most ecologically valid condition tested here. Here target–reference separation and Gabor size (fixed spatial bandwidth) are varied with Gabor spatial frequency (referred to as scaled Gabors). These parametric changes are equivalent to those caused by varying viewing distance. These results are shown in Figs. 2A–C in which stereo sensitivity is plotted as a function of spatial frequency with duration as a parameter, for three subjects.

Stereo sensitivity improves as the Gabor spatial frequency increases for the long (1336 ms) exposure duration, illustrating the well-established spatial frequency/disparity correlation (Hess, Liu, & Wang, 2002; Schor & Wood, 1983). A similar relationship is also seen for the short (122 ms) exposure duration. The slopes of the best fitting lines on these double logarithmic coordinates are significantly steeper for the long exposure (ranging from 0.65 to 1.08 for the long duration and from 0.44 to 0.7 for the short duration—see Table 2).

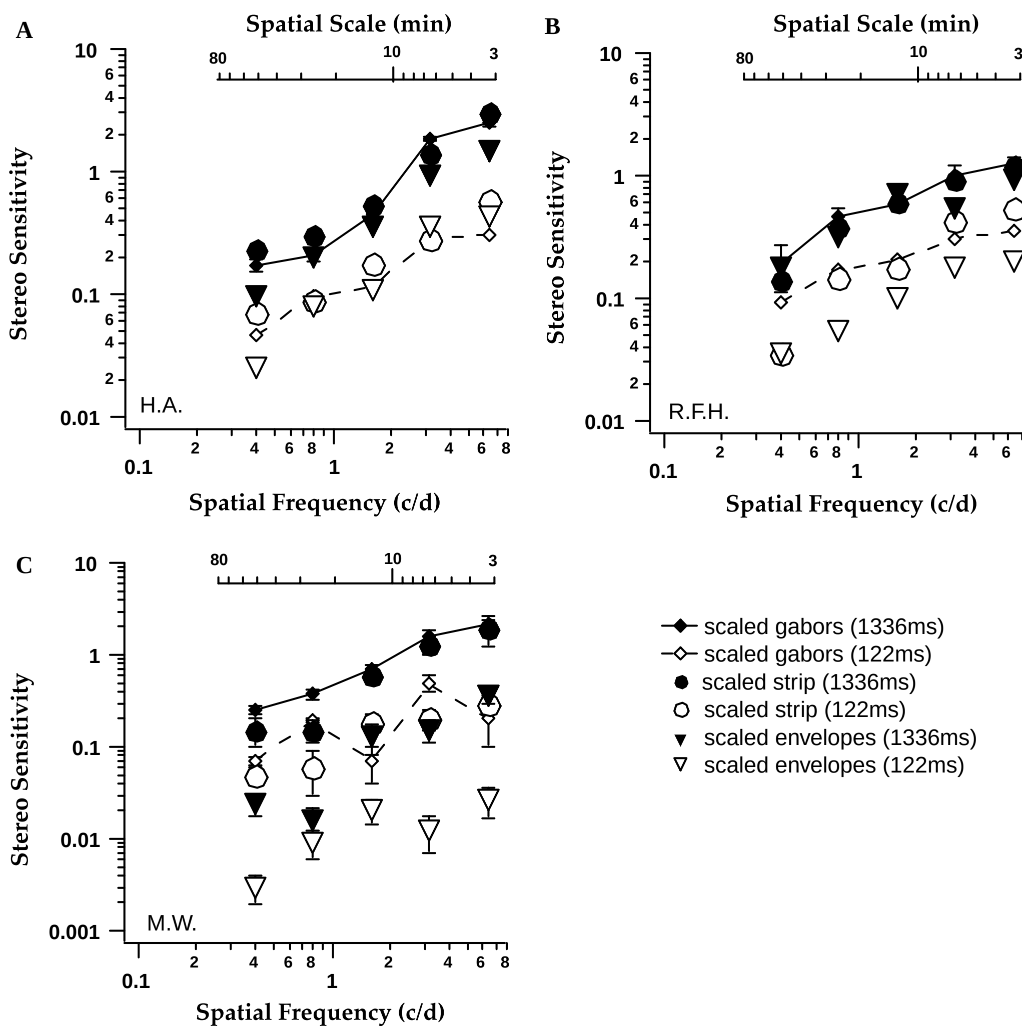

Fig. 2. Stereo sensitivity is plotted as a function of the spatial frequency of the elements for two durations (1336 and 122 ms) for three subjects. Results are compared as a function of spatial frequency for spatially scaled Gabors grating strips and Gaussian envelopes. The top scale refers to the size of the Gaussian sigma of the stimuli.

Table 2
Slopes for spatially scaled stimuli

	Long (1336 ms) duration			Short (122 ms) duration		
	Scaled Gabors	Scaled strips	Scaled envelopes	Scaled Gabors	Scaled strips	Scaled envelopes
HA	1.08 ± 0.1	0.96 ± 0.1	0.99 ± 0.04	0.7 ± 0.1	0.77 ± 0.06	1.04 ± 0.1
RFH	0.65 ± 0.1	0.74 ± 0.1	0.55 ± 0.15	0.47 ± 0.06	0.93 ± 0.1	0.67 ± 0.07
MW	0.83 ± 0.06	1.03 ± 0.1	1.1 ± 0.2	0.44 ± 0.3	0.69 ± 0.1	0.66 ± 0.07

To distinguish between the influences of the Gabor periodicity and the Gabor size, we compared sensitivity for similarly spatially scaled stimuli that were extended horizontally (i.e., a strip of grating—Fig. 1B—referred to as scaled strip) or contained no internal structure (Fig. 1C—referred to as scaled envelopes). In Fig. 2 the results for these two stimuli (scaled strip and scaled envelopes) are compared with those obtained with the scaled Gabors (Fig. 1A). To aid interpretation and comparison of the three sets of results two axes are used; the upper abscissa represents the scale or size of the envelope and the lower abscissa represents spatial frequency. The data obtained using the scaled Gabor stimulus can be interpreted using both axes while the scaled strip stimulus requires the bottom abscissa (spatial frequency) and the scaled envelope stimulus corresponds to the upper abscissa (envelope size).

The results obtained using the scaled strip stimulus were similar to those already described using the scaled Gabor stimulus for both durations, suggesting that the interior structure of the Gabor patch played a dominant role in our first data set. However, comparison of the results obtained using the scaled envelope stimulus shows that although sensitivity is lowered relative to the scaled Gabor stimulus at all sizes and spatial frequencies, it is still possible to perform the task. One subject (HA) exhibits a slightly different pattern, in that his scaled strip and scaled envelope data are more variable and show more overlap than the other subjects, particularly at short exposure durations. These results suggest that he is able to use the envelope under these conditions, and in spite of the variability in his data he, like the others, is most sensitive to the periodic patterns. These data are consistent with previous work by Hess and Wilcox (1994) in showing that stereo sensitivity to Gabor stimuli is determined by both the peak spatial frequency and the envelope size. Our current efforts show that this joint dependence is maintained consistently across a range of target–reference separations, as long as they are scaled with the size and centre frequency of the stimuli.

In the stimuli compared so far, the separation was scaled with the spatial frequency and/or envelope size. To assess the influence of this stimulus attribute we compared sensitivity for scaled Gabors whose separation was held constant (Fig. 1D) with our original measurements using Gabor patterns in which the whole configuration was scaled (Fig. 1A). These results are displayed in Figs. 3A–C for three subjects and show that the effect of separation depends upon the spatial frequency/size of the Gabor and in the case of MW on the exposure duration as well (see Table 3, for slope values). At low spatial frequencies/large envelopes, sensitivity is similar for stimuli of scaled or fixed separation. At higher spatial frequencies/small envelopes, sensitivity is reduced for stimuli of fixed separation.

In the scaled Gabor conditions, and subsequent manipulations, the centre frequency and envelope size of the stimuli covaried. This means that the bandwidth was held constant. It is important to also determine what effect varying the bandwidth has on stereo sensitivity. In this study, we assessed the role of stimulus bandwidth by comparing our initial Gabor results (Fig. 1D) with stimuli in which the Gaussian envelope of the Gabor was fixed, but centre frequency varied—Fig. 1E. Given that the overall size of the stimuli was not varied, we correspondingly held the element separation constant in both conditions. The results are shown in Figs. 4A–C for three subjects. Sensitivity is detrimentally affected if the Gabor bandwidth is too narrow but only at short durations (see Table 4, for slope values). Sensitivity for fixed and variable bandwidth stimuli are similar at the longest duration across the spatial frequency range measured here but sensitivity declines at high spatial frequencies when the duration is short and the bandwidth narrow.

Our results to this point bear upon the relationship between stereo sensitivity and stimulus configuration and provide some indication of the configurations that might best stimulate the neural processes underlying stereopsis. In the natural environment the same object placed at two distances from an observer will exhibit coherent scaling of its surface texture, overall size and relative dimensions; violation of this coherence (non-rigidity) would be rare. Thus when manipulating the frequency content, size and spacing of elements in psychophysical experiments, and measuring sensitivity, it may be important that the spacing be scaled accordingly, while the bandwidth held constant (size is varied along with centre frequency) as shown in Fig. 1A. Reference to the psychophysical literature shows a large number of studies that have used band-limited stimuli such as scaled Gabors or DoGs or even sinewave gratings, and done exactly the opposite: fixed the separation and varied the bandwidth (Harwerth et al., 2003; Schor & Wood, 1983).

While it is well established (Hess et al., 2002; Prince & Eagle, 1999a, 1999b; Schor & Wood, 1983) that frequency, size and bandwidth are all important factors limiting stereo sensitivity in isolation, it is their interaction and combination with spacing that is highlighted here. When considered from this perspective, the size–disparity correlation might best be renamed the "scale–disparity correlation". This is

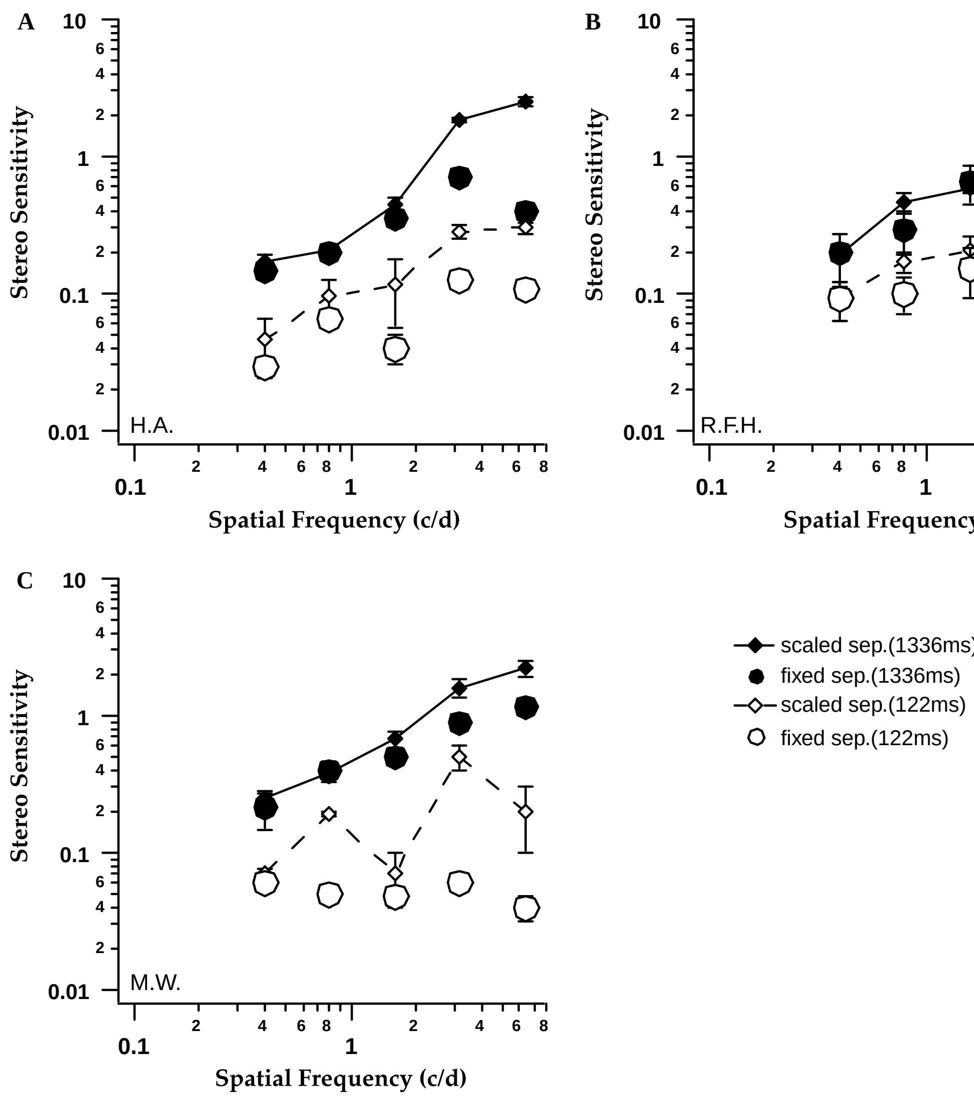

Fig. 3. Stereo sensitivity is plotted as a function of the spatial frequency of the elements for two durations (1336 and 122 ms) for three subjects. Results are compared for stimuli of fixed and scaled element separation. Stimuli with fixed element separations produce poor stereo as the spatial frequency of the stimuli increase.

Table 3
Slopes for stimuli of fixed separation

	Fixed separation	
	Long duration (1336 ms)	Short duration (122 ms)
HA	0.47 ± 0.17	0.48 ± 0.18
RFH	0.42 ± 0.15	0.3 ± 0.04
MW	0.6 ± 0.05	-0.09 ± 0.07

not simply a semantic point, rather, the failure to scale target-stimulus separation in previous experiments (Prince & Eagle, 1999a, 1999b; Prince & Eagle, 2000; Schor & Wood, 1983) may well have resulted in sub-optimal performance in some conditions.

In the preceding study, we have shown that the way stereo *sensitivity* varies with stimulus spatial frequency in turn depends on the particular stimulus configuration used to measure it, with best performance obtained for spatially scaled stimuli. We now assess how the *temporal properties* of stereo processing vary with stimulus spatial frequency and we do this for the same range of stimulus configurations previously described. This allows us to assess whether stereo dynamics are themselves dependent on stimulus configuration, a question that needs to be answered not only for a comprehensive picture of stereo dynamics but also for an adequate comparison with previous studies where different stimulus configurations have been used.

Stereo sensitivity was measured as a function of stimulus duration using a Gaussian temporal envelope with total durations (equivalent to 4× sigma) ranging from 85 to 1336 ms in one-octave steps (7 levels). Thresholds were measured for each of the stimulus configurations (5×5) described above for three subjects. Each temporal test condition was fit with a widely used model of temporal dynamics (see Appendix A for details) based on probability summation of the response of a linear temporal filter (Watson, 1979).

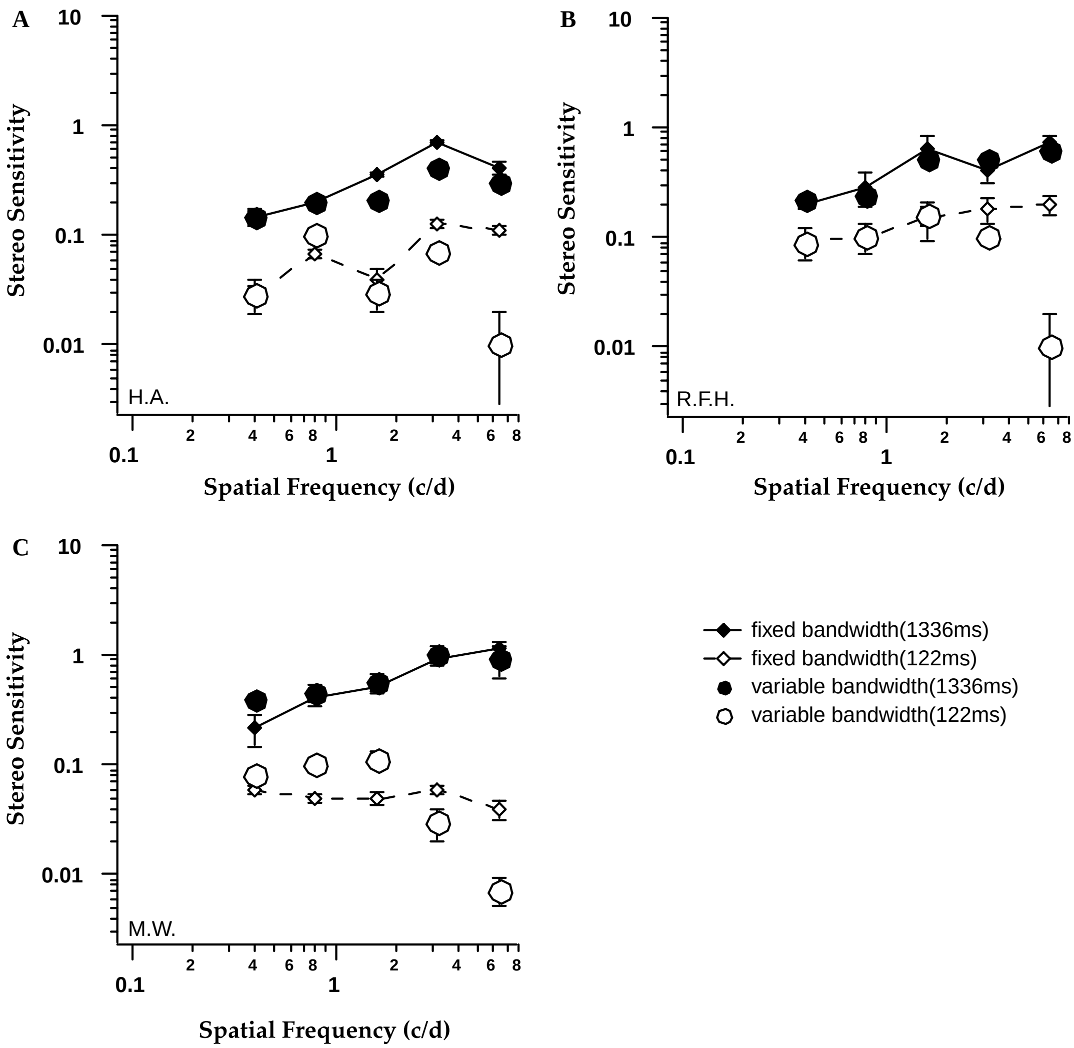

Fig. 4. Stereo sensitivity is plotted as a function of the spatial frequency of the elements for two durations (1336 and 122 ms) for three subjects. Results are compared for stimuli of fixed and variable bandwidth. The combination of high spatial frequency and broad bandwidth results in poor stereo.

Table 4
Slopes for stimuli of variable bandwidth

	Variable bandwidth	
	Long duration (1336 ms)	Short duration (122 ms)
HA	0.3 ± 0.1	-0.35 ± 0.4
RFH	0.4 ± 0.09	-0.6 ± 0.4
MW	0.36 ± 0.07	-0.87 ± 0.3

Fig. 5 displays results for the spatially scaled Gabor configuration (Fig. 5 top left—see Fig. 1A), the spatially scaled strip configuration (Fig. 5 top right—see Fig. 1B) and the spatially scaled envelope configuration (Fig. 5 bottom left—see Fig. 1C) for one subject. The fits to the complete data set were satisfactory giving mean reduced χ^2 values (χ^2 per degree of freedom) of 0.9 ± 0.4 (scaled Gabors), 1.7 ± 0.9 (scaled strips), 1.2 ± 1.0 (scaled envelopes). Fig. 6 displays results for the stimulus of fixed separation and bandwidth (Fig. 6 top—see Fig. 1D) and for the stimulus of fixed separation but variable bandwidth (Fig. 6 bottom—see Fig. 1E) again for one subject. The fits to the complete data set in these conditions were poor giving mean reduced χ^2 values of 7.7 ± 3.7 (fixed sep. constant bandwidth) and 2.8 ± 1.9 (fixed separation, variable bandwidth). In all example cases, stereo sensitivity is plotted against stimulus exposure duration ($4 \times$ Gaussian sigma), for a range of spatial frequencies. The solid and dashed curves are the fits of the temporal integration model to the data (see Appendix A). In all but one case (the highest spatial frequency condition in Fig. 6 bottom, unfilled circles) where the element separation is fixed, the fits are poor.[1] In all cases, stereo sensitivity increases with stimulus duration rapidly up to around 100–150 ms and thereafter more gradually. The form of this relationship did not vary dramatically with stimulus spatial frequency for any of the stimulus configurations except, as already mentioned, for the highest spatial frequency condition in Fig. 6 (bottom, unfilled circles).

[1] Note that this result is consistent with our results from the preceding conditions where we found that when the separation was fixed, performance was poor.

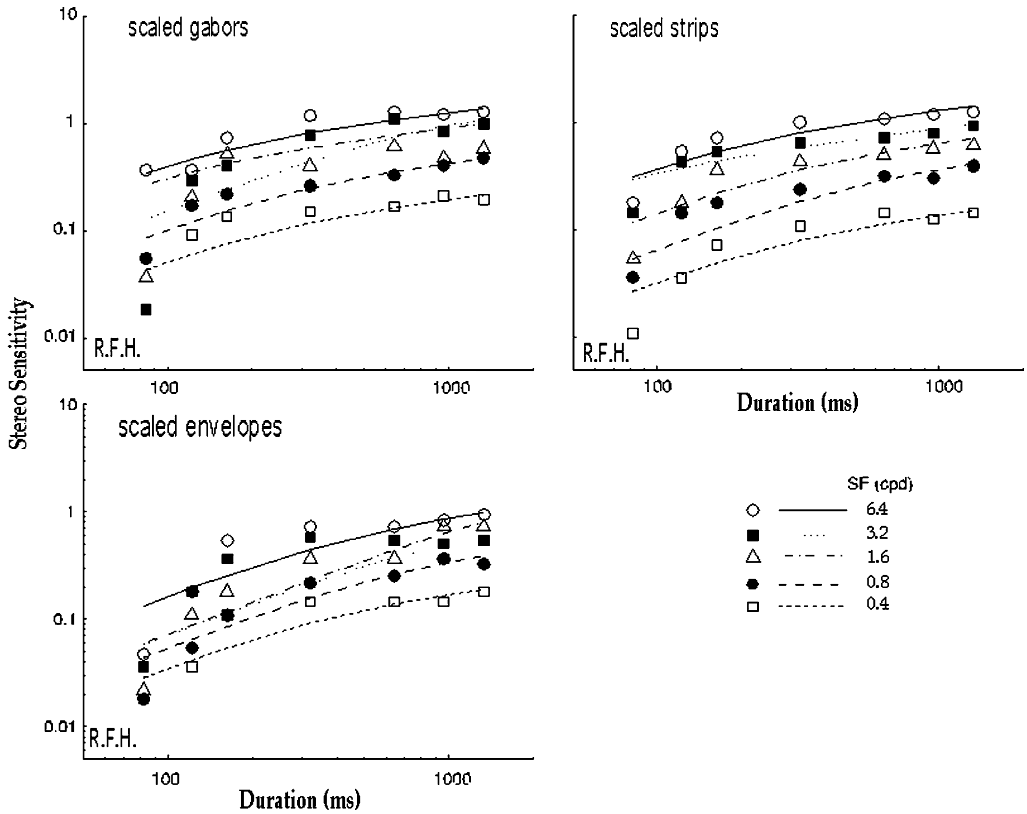

Fig. 5. Stereo sensitivity is plotted against the Gaussian duration (4× sigma) for one subject for three different stimulus configurations (as shown in Figs. 1A–C). Results are shown for five spatial frequencies and the curves are best fits of the temporal summation model detailed in Appendix A.

To obtain a quantitative measure of the relationship between spatial frequency and the dynamics of stereopsis, we used the model fits for each stimulus configuration. The model's parameters were the weighting factors (A, K) and time constants (τ_1 and τ_2) of the two low-pass filters. We did not find a significant difference in the inhibitory component across conditions and subjects: inhibition was weak and slow in all cases. This suggests a weak transient component compared to the sustained one for accounting for the dependence of sensitivity on stimulus duration. Only the weighting factor, A and time constant, τ_1 of the excitatory components changed across conditions and subjects. Therefore we focused our analysis on τ_1, which characterizes the sustained dynamics of the processing associated with the stereo task. Fig. 7 shows scatterplots for the (τ_1, spatial frequency) pairs derived from the fitting procedure for each experimental condition, along with a 95% confidence ellipse on their log transformation. Each ellipse encloses 95% of the data points (assuming the two variables have a bivariate normal distribution), indicating the area in the plot that accounts for the vast majority of the data. The major axis of each ellipse is the straight line regression of time constant on spatial frequency (on log scale). The centre of the ellipse is the mean of time constants and spatial frequencies and the ratio of the major to minor axes is the ratio of the standard deviations of the data points' distances from the centre, projected onto the major and minor axes. There is a great deal of variability in this parameter across subjects and conditions but no systematic variation with spatial frequency for any stimulus configuration. This analysis confirms our observations from the raw data, shown in Fig. 5, that stereo dynamics are sustained and do *not* vary systematically with spatial scale.

As is evident in Figs. 5 and 6, the model did not accurately capture performance at the shortest exposure durations. In order to assess whether there might be a strong spatial frequency/spatial scale dependent response at the shortest durations that the model missed, we undertook a separate analysis in which we compared the averaged threshold loss for the two shortest durations as a function of spatial frequency for the five stimulus configurations used. These results are shown in Fig. 8. No clear relationship was found between the sensitivity reduction at the two shortest exposure durations and stimulus spatial frequency (or stimulus spatial scale in terms of the scaled envelope configuration) for any of our stimulus configurations. For example, if the response at low spatial frequencies is more transient than that at high spatial frequencies we would have expected to see less reduction in sensitivity at low versus high spatial frequencies (i.e., a strong positive relationship) which was not observed.

5. Discussion

The main finding from this study is that the dynamics of stereo processing do not vary systematically with spatial frequency/spatial scale. Since contrast detection dynamics

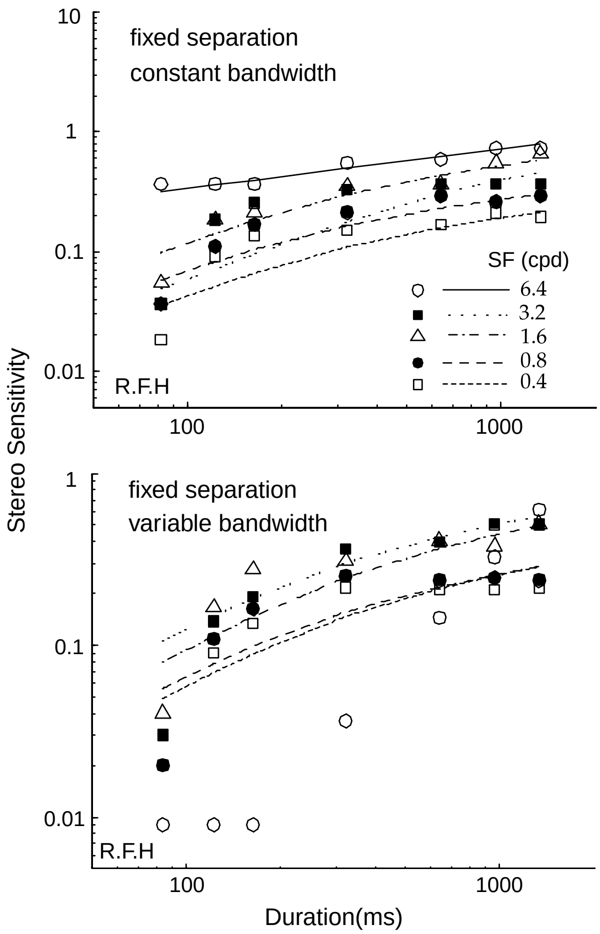

Fig. 6. Stereo sensitivity is plotted against the Gaussian duration (4× sigma) for one subject for two different stimulus configurations (as shown in Figs. 1D and E). Results are shown for five spatial frequencies and the curves are best fits of the temporal summation model detailed in Appendix A. The data obtained with the fixed separation, variable bandwidth condition was not able to be fit by the model.

do vary with spatial frequency, with low frequencies being processed more quickly than high spatial frequencies, the expectation here was for faster dynamics for low spatial frequency stereo processing. This was not the case. This finding is in agreement with a recent study where a different methodological approach to the assessment temporal processing was used, namely measurement of the critical duration (Harwerth et al., 2003). In their study, a similarly high dispersion of values was shown for critical duration but with no systematic variation with stimulus spatial frequency. Harwerth et al. (2003) used stimuli of fixed spatial bandwidth but fixed separation. We show that while the absolute sensitivity of stereo can depend on stimulus configuration, the relationship between the dynamics and spatial frequency do not.

This outcome has important implications for models of human stereopsis; as it bears upon a common assumption implicit in the coarse-to-fine model of stereo processing which has attracted computational (Marr & Poggio, 1979; Nishihara, 1984; Quam, 1987), psychophysical (Rohaly & Wilson, 1993; Watt, 1987; Wilson, Blake, & Halpern, 1991), (but also see Jones & Malik, 1992); (Smallman & MacLeod, 1995) as well as neurophysiological (Menz & Freeman, 2003, 2004a, 2004b) support. That is, that the directionality of processing is evident in the temporal dynamics of stereopsis, with the disparities represented by low spatial frequency, large stimuli being processed before those represented by high spatial frequencies, small stimuli. The present evidence, along with that of Harwerth et al. (2003), suggests that the speed at which disparities at different scales are processed is the same. Therefore, to maintain the coarse-to-fine model, the scale-dependent dynamics must occur, not as the neurophysiology suggests, at the level of disparity coding, but at the level of scale combination. At such a later site, the dynamics of scale combination need not be rigidly set and could operate in either coarse-to-fine (Rohaly & Wilson, 1993; Watt, 1987; Wilson et al., 1991) or fine-to-coarse modes (Jones & Malik, 1992; Smallman & MacLeod, 1995).

A difference exists between contrast thresholds that do exhibit scale-dependent dynamics and stereo thresholds that do not. Our use of stimuli of comparable suprathreshold contrast ensured that stereo performance was not affected by reduced stimulus visibility as presentation duration was shortened. Furthermore, the particular suprathreshold level that we chose to compare stimuli (i.e., 3.5× threshold) is unlikely, in itself, to have affected our conclusions, as previous studies have shown that the contrast of stimuli, as long as they are suprathreshold, does not affect the dynamics (Harwerth et al., 2003; Ogle & Weil, 1958).

5.1. Stimulus configuration

The configuration effects reported here show that some stimulus arrangements produce better stereo sensitivity than others. Our study of this is one of the first to have been done and it has long been an unaddressed issue. However, in retrospect, the results are not that surprising. That is, the optimal arrangement when varying spatial frequency is to scale proportionally all other stimulus spatial dimensions (i.e. separation and size of Gabors). Non-optimal stereo sensitivity is found at higher spatial frequencies when either separation or Gabor size is not scaled with spatial frequency. In the case of separation, it is a consequence of either the greater fall-off in high spatial frequency stereo sensitivity with eccentricity (Schor & Badcock, 1985) or a smaller critical distance over which high spatial frequency stereo comparisons can be made. The finding that bandwidth is an important consideration here suggests that a matching problem might limits performance because of both the distance between features and the number of false matches within a feature (Prince & Eagle, 2000; Ziegler, Kingdom, & Hess, 2000). The finding that short durations appear to exacerbate this problem is a novel and unexpected finding. It is possible that the additional viewing time permits vergence eye movements that help resolve the matching ambiguity. More research will be needed to determine if this phenomenon has a neural or oculomotor basis.

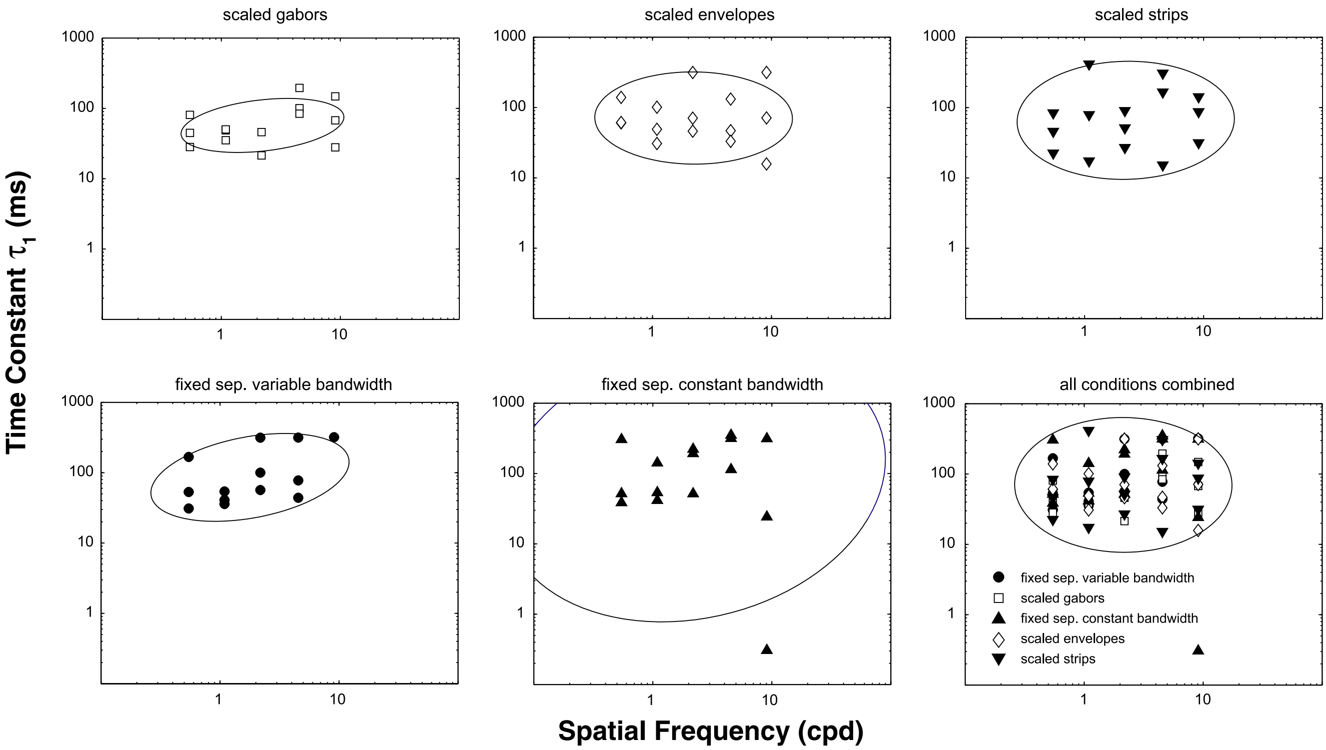

Fig. 7. Summary scatter plot of the time constant derived from the model fits to the data as a function of stimulus spatial frequency for all three subjects under all stimulus configurations (5) and spatial frequencies (5). The ellipse defines the 95% confidence limits. There is no systematic variation in the dynamics with stimulus spatial frequency.

5.2. The size–disparity correlation

There are currently two general classes of neurophysiological models of disparity coding. One depends on the activation of spatially coincident quadrature-phase sensitive detectors (Ohzawa, De Angeles, & Freeman, 1996) and the other depends upon the activation of spatially disparate but phase-aligned detectors (Barlow, Blakemore, & Pettigrew, 1967). The former, 'phase-coding' model, requires a linear relationship between stimulus spatial frequency and stereo sensitivity whereas the latter, 'position-coding model' does not. Such a linear relationship has been reported by Schor and Wood (1983) for difference-of Gaussian stimuli whose overall dimensions scale with the peak spatial frequency of the elements, by Smallman and MacLeod (1994) for bandpass filtered random dot stereograms, by Prince and Eagle (1999a, 1999b) for Gabors of fixed bandwidth and by Hess et al. (2002) for fractal noise discs. The present results suggest: (1) that the slope of this relationship is not fixed but depends on stimulus variables such as duration, separation, bandwidth and (2) a more pertinent correlation is between stereo sensitivity and spatial scale (i.e. the combination of spatial frequency, size and spacing) rather than stereo sensitivity and spatial frequency per se.

5.3. Sustained versus transient dichotomy

Recently the stereoscopic system has been described as having separate transient and sustained mechanisms (Edwards, Pope, & Schor, 1999, 2000; Edwards & Schor, 1999; Pope, Edwards, & Schor, 1999a, Pope, Edwards, & Schor, 1999b; Schor et al., 1998). In their model of the temporal properties of stereopsis, Cormack and Landers (1997) demonstrated that it is not necessary to posit two distinct channels, a sustained channel and a transient channel. The summary graph depicted in Fig. 7 support this position, as they do not show any obvious dichotomy. The raw data show that there is a monotonic improvement in sensitivity with increasing exposure duration that is a characteristic of a sustained system. The modeling showed that the inhibitory component was very weak compared with the sustained component, irrespective of stimulus spatial frequency, size or spacing. It is possible that the experiments of Schor and his colleagues (Edwards et al., 1999, Edwards, Pope, & Schor, 2000; Edwards & Schor, 1999; Pope et al., 1999a, 1999b; Schor et al., 1998) map onto another stimulus attribute, namely whether the stimulus is defined by luminance or contrast variations. The stimuli used here did not specifically target either 1st or 2nd order processes, but given the presence of suprathreshold luminance-based disparity signals it is likely that our stimuli activated the high-resolution luminance-based system (Hess & Wilcox, 1994).

In conclusion, the experiments reported here first evaluate the effects of stimulus configuration on *stereo sensitivity* and showed that test–reference spacing must be scaled with stimulus size and frequency in order to attain

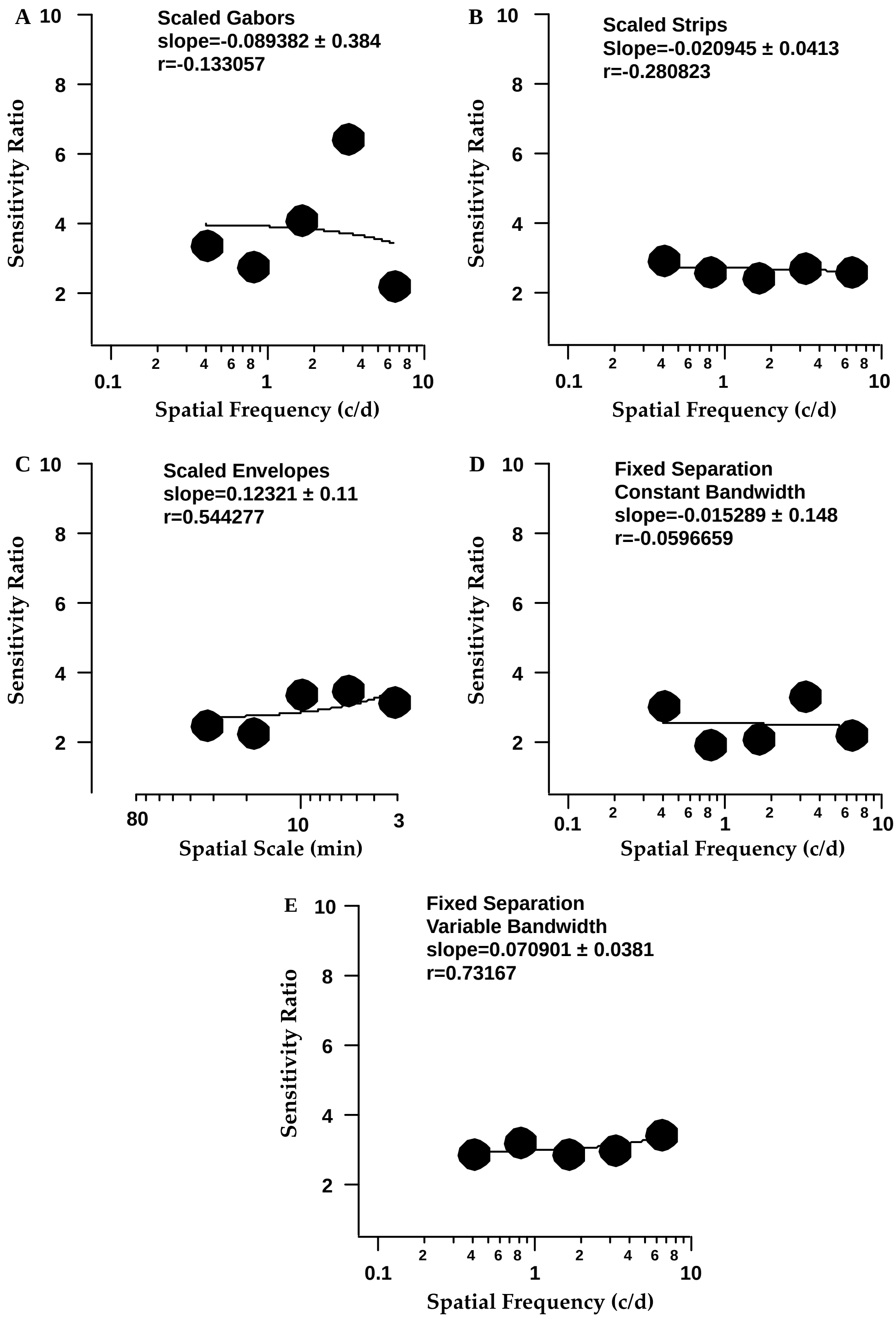

Fig. 8. An analysis of the sensitivity loss produced by the two shortest exposure durations expressed as a ratio and plotted against stimulus spatial frequency/spatial scale for the five stimulus configurations used, ((A)–(E), see Fig. 1). The slope of the best fitting linear function and the Pearson's *r* value is given for each. The loss of sensitivity at short durations was not strongly dependent on spatial frequency/spatial scale.

good stereo sensitivity. Second, the *dynamics* of stereo processing, unlike those of contrast detection, do not vary systematically with stimulus spatial frequency regardless of stimulus configuration. Third, for band-limited stimuli with a wide range of configurations stereo sensitivity is temporally sustained with no evidence for a transient/sustained dichotomy.

Appendix A. Temporal integration model

The sensitivity data were fitted to a model based on probability summation over time of the response of a linear temporal filter (Watson, 1979). According to this concept, the model sensitivity is given by the Quick formula:

$$S = \left[\int |R(t)|^{\beta} \mathrm{d}t \right]^{1/\beta},$$

where b may be derived from the slope of the psychometric function or may be adjusted to fit the data; $R(t)$, the visual response to a given stimulus, is given by the convolution of the detection filter's impulse response $I(t)$ with the temporal envelope of the stimulus $G(t)$:

$$R(t) = \int I(\tau) \cdot G(t-\tau)\mathrm{d}\tau.$$

The linear temporal filter has an impulse response $I(t)$ with excitatory and inhibitory components, each approximated by a cascaded low-pass leaky integrator (Watson, 1986):

$$I(t) = A \cdot \Big\{ \varepsilon(t)[\tau_1(n_1-1)!]^{-1}(t/\tau_1)^{n_1-1}\exp(-t/\tau_1) - K \cdot \varepsilon(t)[\tau_2(n_2-1)!]^{-1}(t/\tau_2)^{n_2-1}\exp(-t/\tau_2) \Big\},$$

where $e(t)$ is the step function, t_1 and t_2 the time constants of the two components n_1 and n_2 are the number of cascaded low-pass stages of each component, A is a sensitivity factor and K is the weighting factor of the inhibitory component.

All stimuli have a Gaussian temporal envelope $G(t)$:

$$G(t) = \exp\left[-0.5\left(\frac{t-t_0}{\sigma_t} \right)^2 \right],$$

where t_0 is $0.25 \times$ the presentation duration at which the stimuli have their highest contrast (3.5× threshold), and s_t the time constant of the stimulus.

This model was fitted to the sensitivity data S plotted as function of s_t. We retained A, K, t_1, and t_2 as free parameters while fixing the other model's parameters ($n_1 = 9$, $n_2 = 10$, $b = 3.0$). We used Matlab (The MathWorks) to fit the model to each sensitivity dataset using a least squares weighted procedure and the Nelder–Mead simplex optimization. Each fit was repeated at least 25 times with random initial parameters to verify the stability of the solution, and the final estimates of the parameters were provided by the best fit.

References

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). The neural mechanism of binocular depth discrimination. *Journal of Physiology (London), 193*, 327–342.

Breitmeyer, B., & Ogmen, H. (2000). Recent models and findings in visual backward masking:A comparison, review and update. *Perception and Psychophysics, 62*, 1572–1595.

Burr, D. C. (1981). Temporal summation of moving images by the human visual system. *Proceedings of the Royal Society of London, Series B, 211*, 321–339.

Cormack, L. K., & Landers, D. D. (1997). Some spatio-temporal interactions in stereopsis. *Investigative Ophthalmology & Visual Science (Suppl), 38*(4), p4253.

Edwards, M., Pope, D. R., & Schor, C. M. (1999). Orientation tuning of the transient-stereopsis system. *Vision Research, 39*(16), 2717–2727.

Edwards, M., Pope, D. R., & Schor, C. M. (2000). First- and second-order processing in transient stereopsis. *Vision Research, 40*(19), 2645–2651.

Edwards, M., & Schor, C. M. (1999). Depth aliasing by the transient-stereopsis system. *Vision Research, 39*(26), 4333–4340.

Harwerth, R. S., Fredenburg, P. M., & Smith, E. L. 3rd, (2003). Temporal integration for stereoscopic vision. *Vision Research, 43*(5), 505–517.

Hess, R. F., Liu, H. C., & Wang, Y-Z. (2002). Luminance spatial scale and local stereo-sensitivity. *Vision Research, 42*, 331–342.

Hess, R. F., & Wilcox, L. M. (1994). Linear and non-linear filtering in stereopsis. *Vision Research, 34*, 2431–2438.

Jones, D. G., & Malik, J. (1992). Computational framework for determining stereo correspondence from a set of linear spatial filters. *Image and Vision Computing, 10*, 699–708.

Julesz, B. (1964). Binocular depth perception without familiarity cues. *Science, 145*, 356–362.

Kahneman, D. (1968). Methods findings and theory in studies of visual masking. *Psychological Bulletin, 70*, 404–425.

Liss, P. (1968). Does backward masking by visual noise stop visual processing? *Perception and Psychophysics, 4*, 328–340.

Mallot, H. A., Gillner, S., & Arndt, P. A. (1996). Is correspondence search in human stereo vision a coarse-to-fine process? *Biological Cybernetics, 74*, 95–106.

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society, Series B, 204*, 301–328.

Menz, M. D., & Freeman, R. D. (2003). Stereoscopic depth processing in the visual cortex: a coarse-to-fine mechanism. *Nature Neuroscience, 6*(1), 59–65.

Menz, M. D., & Freeman, R. D. (2004a). Functional connectivity of disparity-tuned neurons in the visual cortex. *Journal of Neurophysiology, 91*(4), 1794–1807.

Menz, M. D., & Freeman, R. D. (2004b). Temporal dynamics of binocular disparity processing in the central visual pathway. *Journal of Neurophysiology, 91*(4), 1782–1793.

Nishihara, H. K. (1984). Practical real time image stereo matcher. *Optical Engineering, 23*, 536–545.

Ogle, K. N., & Weil, M. P. (1958). Stereoscopic vision and the duration of the stimulus. *A.M.A. Archives of Ophthalmology, 59*, 4–17.

Ohzawa, I., De Angeles, G. C., & Freeman, R. D. (1996). Encoding of binocular disparity by simple cells in cat's visual cortex. *Journal of Neurophysiology, 75*(5), 1779–1805.

Pope, D. R., Edwards, M., & Schor, C. M. (1999a). Orientation and luminance polarity tuning of the transient-vergence system. *Vision Research, 39*(3), 575–584.

Pope, D. R., Edwards, M., & Schor, C. M. (1999b). Extraction of depth from opposite-contrast stimuli: transient system can, sustained system can't. *Vision Research, 39*(24), 4010–4017.

Prince, S. J., & Eagle, R. A. (1999a). Size–disparity correlation in human binocular depth perception. *Proceedings of the Royal Society of London, Series B: Biological Sciences, 266*(1426), 1361–1365.

Prince, S. J. D., & Eagle, R. A. (1999b). Size–disparity correlation in human binocular depth perception. *Proceedings of the Royal Society of London, Series B, 266*, 1361–1365.

Prince, S. J., & Eagle, R. A. (2000). Stereo correspondence in one-dimensional Gabor stimuli. *Vision Research, 40*(8), 913–924.

Quam, L. H. (1987). Hierarchical warp stereo. In M. A. Fischler, & O. Firschein (Eds.), *Readings in computer vision* (pp. 80–86). Los Altos: Kauffman.

Rohaly, A. M., & Wilson, H. R. (1993). The nature of coarse-to-fine constraints on binocular fusion. *Journal of the Optical Society of America A, 10*, 2433–2441.

Rohaly, A. M., & Wilson, H. R. (1994). *Vision Research, 34*, 1315–1325.

Schor, C. M., & Badcock, D. R. (1985). A comparison of stereo and vernier acuity within spatial channels as a function of distance from fixation. *Vision Research, 25*, 1113–1119.

Schor, C. M., Edwards, M., & Pope, R. D. (1998). Spatial-frequency and contrast tuning of the transient-stereopsis system. *Vision Research, 38*(20), 3057–3068.

Schor, C. M., & Wood, I. (1983). Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision Research, 23*, 1649–1654.

Smallman, H. S., & MacLeod, D. I. A. (1994). Size–disparity correlation in stereopsis at contrast threshold. *Journal of the Optical Society of America A, 11*, 2169–2183.

Smallman, H. S., & MacLeod, D. I. A. (1995). Fine-to-coarse scale disambiguation in stereopsis. *Vision Research, 35*, 1047–1060.

Sperling, G. (1963). A model for visual memory tasks. *Human Factors, 5*, 19–31.

Uttal, W. R., Fitzgerald, J., & Eskin, T. E. (1975). Parameters of tachistoscopic stereopsis. *Vision Research, 15*, 705–712.

Watson, A. B. (1979). Probability summation over time. *Vision Research, 19*(5), 515–522.

Watson, A. B. (1986). Temporal sensitivity. In L. K. K. R. Boff & J. P. Thomas (Eds.), *Handbook of perception and human performance I: Sensory processes and perception* (pp. 6.1–6.41). New York: John Wiley and Sons.

Watt, R. J. (1987). Scanning from coarse to fine spatial-scales in the human visual system after the onset of the stimulus. *Journal of the Optical Society of America, 4*, 2006–2021.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics, 63*(8), 1314–1329.

Wilson, H. R., Blake, R., & Halpern, D. L. (1991). Coarse spatial scales constrain the range of binocular fusion on fine scales. *Journal of the Optical Society of America A, 8*, 229–236.

Ziegler, L. R., Kingdom, F. A., & Hess, R. F. (2000). Local luminance factors that determine the maximum disparity for seeing cyclopean surface shape. *Vision Research, 40*(9), 1157–1165.